Computer Program of Forward Selection and Backward Elimination Procedure in Multiple Regression Analysis

Hirokazu OSAKI* and Susumu KIKUCHI*

(Received December, 10, 1974)

Synopsis

Multiple regression analysis are often used to explain the relation between the dependent variable and the independent variables. In case of that it arises necessity that the important independent variables which are closely correlated with the dependent variable are selected from among all given ones. There are some selection procedures. But these procedures can't be used usefully without using computer.

Therefore two selection procedures that is Forward selection procedure and Backward elimination procedure in multiple regression analysis are programmed by Fortran IV.

1. Introduction

It is important in multiple regression analysis to select or order the independent variables which are closely correlated with the dependent variable in order to explain the phenomenon clearly. There are some procedures of selecting the important independent variables (1,2,3). But these procedures can be applied to the various problems only by using of computer.

In this paper we mention on the Fortran IV program of Forward
selection procedure and Backward elimination procedure in multiple regression analysis.

2. Analytical method

Set of independent variables is put as follows.

\[ C = \{x_1, x_2, \ldots, x_k\} \]

k: number of independent variables

Y is the dependent variable.

\( (y_i, x_{i1}, x_{i2}, \ldots, x_{ik}) \), \( i=1,2,\ldots,n \) are the data of Y and C.

The mean square due to residual variation \( S_y^2 \) is one scale of precision of estimate of the regression line in using independent variables.

\[
S_y^2 = \frac{\sum_{m=1}^{n} (y_m - \bar{y})^2 - d' S^{-1} d}{(n - k - 1)} \quad (1)
\]

where \( d' = (d_1, d_2, \ldots, d_k)' \): k x 1 vector

\[ d_i = \sum_{m=1}^{n} (y_m - \bar{y})(x_{im} - \bar{x}_i) \]

\[ S = (s_{ij}) \text{ : k x k matrix} \]

\[ s_{ij} = \sum_{m=1}^{n} (x_{im} - \bar{x}_i)(x_{jm} - \bar{x}_j) \]

The effect of independent variables to the dependent variable is contained in \( d' S^{-1} d \). But it is necessary to evaluate the effect of each independent variable or that of subset of independent variables in order to explain clearly the relation between independent variables and dependent one. There are some criterions of evaluating the effect. Therefore the procedures which select or order the independent variables have been proposed by many authors(1,2,3).

In this paper we deal with the procedures that the independent variable is selected or eliminated one by one from among all given independent variables in accordance with the following mentioned criterions. These procedures are called Forward selection procedure and Backward elimination procedure(3).
D(i_1, i_2, \ldots, i_m) is put as the value of \( d'S^{-1}d \) which is calculated from independent variables \( \{x_{i_1}, x_{i_2}, \ldots, x_{i_m}\} \).

E(m, i_p) is put as the value of \( d'S^{-1}d \) which is calculated from independent variables \( \{x_{i_1}, x_{i_2}, \ldots, x_{i_m}\} - \{x_{i_p}\} \).

2.1 Forward selection procedure

\( x_{i_m} \) is put as the independent variable which satisfies the following relation.

\[
\max_{i_m} D(I_1, I_2, \ldots, I_{m-1}, i_m) \quad \{x_{i_m}\} \subset C - \{x_{I_1}, x_{I_2}, \ldots, x_{I_{m-1}}\} \quad \ldots \ldots (2)
\]

Then the locally best regression line and the mean square due to residual variation are as follows.

\[
\hat{\gamma} = \sum_{j=1}^{m} a_{ij}(x_{ij} - \bar{x}_{ij}) + \bar{y}
\]

\[
S_{ym}^{2} = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2 - D(I_1, I_2, \ldots, I_m)}{(n - m - 1)}
\]

Where \( m = 1, 2, 3, \ldots, k \).

2.2 Backward elimination procedure

\( x_{j_{k-m+1}} \) is put as the independent variable which satisfies the following relation.

\[
\max_{i_p} E(k-m+1, i_p) \quad \{x_{i_p}\} \subset C - \{x_{j_k}, x_{j_{k-1}}, \ldots, x_{j_{k-m+2}}\} \quad \ldots \ldots (3)
\]

Then the locally best regression line and the mean square due to residual variation are as follows.

\[
\hat{\gamma} = \sum_{j=1}^{k-m} b_{ij}(x_{ij} - \bar{x}_{ij}) + \bar{y}
\]

\[
S_{yk-m}^{2} = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2 - E(k-m+1, j_{k-m+1})}{(n-k+m-1)}
\]
Where \( m = 1,2,\ldots,k-1 \). And for \( m=1 \), \( \{x_{j_k}, x_{j_{k-1}}, \ldots, x_{j_{k-m+2}}\} = \emptyset \), and for \( m=k-1 \), \( x_{j_1} = C - \{x_{j_k}, x_{j_{k-1}}, \ldots, x_{j_2}\} \).

These two procedures have strong and weak points. But if two procedures are used at the same time, the each strong points can partially make up for each weak point. Therefore these two procedures can be programmed in one program.

3. Program

This program is written by Fortran IV and is the form of subroutine (4). The subroutine name is FORBAC.

```
SUBROUTINE FORBAC(AAA,AA1,MA,KKK,NKK,STORE)
```

3.1 Argument List

<table>
<thead>
<tr>
<th>ARGUMENT</th>
<th>I/O</th>
<th>TYPE</th>
<th>SIZE</th>
<th>DEFINITION</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAA</td>
<td>INPUT</td>
<td>REAL</td>
<td>50 x 50</td>
<td>unbiased variance covariance matrix</td>
</tr>
<tr>
<td>AA1</td>
<td>INPUT</td>
<td>REAL</td>
<td>50</td>
<td>mean vector</td>
</tr>
<tr>
<td>MA</td>
<td>INPUT</td>
<td>INTEGER</td>
<td>1</td>
<td>number of data</td>
</tr>
<tr>
<td>KKK</td>
<td>INPUT</td>
<td>INTEGER</td>
<td>50</td>
<td>dependent and independent variables number</td>
</tr>
<tr>
<td>NKK</td>
<td>INPUT</td>
<td>INTEGER</td>
<td>1</td>
<td>number of independent variables + 1</td>
</tr>
<tr>
<td>STORE</td>
<td>OUTPUT</td>
<td>REAL</td>
<td>50 x 4</td>
<td>results of two procedures</td>
</tr>
</tbody>
</table>

3.2 Suggestions on using

3.2.1 \( NKK \leq 50 \)

3.2.2 If for some \( i \), \( AAA(i,i) = 0 \), then the computation stops as \( i \)-th column is used in the calculation.

3.2.3 Correspondence between arguments and given data

\[
AAA(i,j) = a_{ij}, \quad a_{11} = \sum_{m=1}^{n} (y_m - y)^2 / (n - 1)
\]
aa_{i+1} = aa_{i+1} = d_i / (n - 1), i,j = 1, 2, ..., k

\[ \text{AA}_1(i) = c_1, \quad c_1 = \overline{Y}, \quad c_{i+1} = \overline{X}_i, \quad i=1, 2, ..., k \]

\[ MA = n, \quad n : \text{number of data} \]

\[ \text{KKK}(i) = k_i, \quad k_i : \text{dependent variable number} \]

\[ k_{i+1} = \text{independent variable number} \]

\[ NKK = k + 1, \quad k : \text{number of independent variables} \]

\[ \text{STORE}(i,j) = \text{st}_{ij} \]

\[ \text{st}_{i1} = I_i \]

\[ \text{st}_{i2} = D(I_1, I_2, ..., I_i) \]

\[ \text{st}_{i3} = J_i \]

\[ \text{st}_{i4} = E(i+1, J_i + 1) \]

3.2.4 Subroutine PRINTA and SIMEQS are used in FORBAC. PRINTA is used to print out the results. And SIMEQS is used to solve the linear equation. These subroutine are listed in FORBAC. Program listing is shown in Table 1.

4. Example

The data in a four variable (k = 4) problem given by A. Hald(5) are used to check the program. This data were used by N.R.Draper and H.Smith(3) too. Given data are shown in Table 2. And results are shown in Table 3.

References

(2) P.S.Dwyer "Linear computations", John Wiley & Sons, Inc., (1960)
Table 1. Program Listing

SUBROUTINE FORBAC(AA,AI,MA,KK,NK,STORE)

SUBROUTINE OF FORWARD AND BACKWARD SELECTION OF VARIABLES

IN MULTIPLE REGRESSION

AAA(50,50) VARIANCE COVARIANCE MATRIX

AA(50) MEAN VECTOR

MA NUMBER OF DATA

KK(50) DESIGNATED NUMBER OF VARIABLES

NK NUMBER OF DESIGNATED NUMBER OF VARIABLES

STORE(50,4) RESULTS OF SELECTION PROCEDURES

DIMENSION AAA(50,50),AA(50),KK(50),NJSTOR(50),STORE(50,4),
1

KMOQ(50,4),KKK(50),CC(50,50),DIFF(50),
2

STAND(50),SSSTAND(50),WORK(50,2),BB(50)

WRITE(6,3000)

3000 FORMAT(1H1,///,20X,** ORDERING OF VARIABLES IN MULTIPLE **,
1

1 REGRESSION **,** ,///)

WRITE(6,3005) KK(1)

3005 FORMAT(1H ,20X,'DEPENDENT VARIABLE NUMBER = ,I5,/) 

NK=NKK-1

YMEAN=AA(I)

YVAR=AAA(I)

DO 22 I=1,NKK

22 KKK(I)=KK(I)

WRITE(6,3010)

3010 FORMAT(1H ,20X,'INDEPENDENT VARIABLE NUMBER = ,I5,/) 

WRITE(6,3020) (KKK(I),I=1,NK)

3020 FORMAT(1H ,40X,I6,20X,E15.7)

12 CONTINUE

2 CONTINUE

DO 20 I=1,NKK

20 CONTINUE

DO 21 J=1,NKK

21 CONTINUE

WRITE(6,3040)

3040 FORMAT(1H ,///,20X,'NORMALIZED COEFFICIENTS',///,
1

20X,'INDEPENDENT VARIABLE NUMBER NORMALIZED COEFFICIENT') 

DO 25 I=1,NK

25 WRITE(6,3050) KKK(I),DIFF(I)

3050 FORMAT(1H ,30X,I6,20X,E15.7)
WRITE(6,3060)
3060 FORMAT(1H1,/../20X,'CORRELATION MATRIX'/../20X
1 'BETWEEN INDEPENDENT VARIABLES CORRELATION COEFFICIENT')
DO 26 I=1,NKK
DO 26 J=1,NKK
26 WRITE(6,3065) KKK(I),KKK(J),CCC(I,J)
3065 FORMAT(1H1,25X,'(',15f2,',',15f2,')',15X,E15.7)

C C
C FORWARD SELECTION PROCEDURE
C
WRITE(6,3070)
3070 FORMAT(1H1,/../30X,'** FORWARD SELECTION PROCEDURE **',/../)
DO 100 I=1,NKK
100 KKMOD(I,1)=1
J=1
AMAX=0.0
DO 110 J=1,NKK
ASQR=DIFF(I)**2
IF(ASQR<AMAX) 110,110,120
110 MMAX=I
AMAX=ASQR
120 CONTINUE

KKMOD(J,2)=MMAX
KKMOD(MMAX,1)=0
WORK(1,1)=AMAX
STORE(1,1)=KKMOD(1,2)
STORE(1,2)=AMAX
STAND(I)=DIFF(MMAX)/SSTAND(MMAX)
KKK(J,1)=MMAX
CALL PRINTA(I,MMAX,MA,AMAX,KKK,KKK,STAND,AA1,1,NKK,YVAR,YMEAN)
IF(NKK.EQ.1) GO TO 110
DO 130 J=2,NKK
130 IF(AMAX.EQ.0) GO TO 120
DO 140 KK=1,NKK
140 CONTINUE
AAA(L,M)=CCC(KP,KQ)
AAA(L,J)=CCC(KP,KR)
AAA(J,L)=AAA(L,J)
AA1(L)=DIFF(KP)
BB1(L)=AA1(L)
150 CONTINUE
AAA(L,J)=CCC(KP,KR)
AA1(L)=DIFF(KP)
BB1(J)=AA1(J)
CALL SIMEQ(AAA,AA1,J,NCHEC)
IF(NCHEC.NE.1) GO TO 162
161 WRITE(6,3081)
3081 FORMAT(1H1,/../20X,'DIAGONAL ELEMENT OF MATRIX IS ZERO')
RETURN
162 CONTINUE
ASQR=0.0
DO 160 I=1,J
160 CONTINUE
ASQR = ASQR + AA1(I) * BB1(I)
IF (ASQR = AMAX) 180, 180, 170
AMAX = ASQR
MHAN = KK
DO 175 I = 1, J
175 STAND(I) = AA1(I)
180 CONTINUE
140 CONTINUE
KKMOD(J, 2) = MHAN
KKMOD(MHAN + 1) = 0
WORK(J, 1) = AMAX
STORE(J, 1) = KKMOD(J, 2)
STORE(J, 2) = WORK(J, 1)
DO 9510 LL = 1, NKK
9510 AA1(LL) = WORK(LL, 2)
DO 9520 LL = 1, J
KKKK(LL) = KKMOD(LL, 2)
LN = KKK(L)
STAND(LL) = STAND(LL) / SSTAND(LN)
9520 CONTINUE
CALL PRINTA(J, MHAN, MA, AMAX, KKK, KKKK, STAND, AA1, 1, NKK, YVAR, YMEAN)
130 CONTINUE
STORE(NKK, 4) = STORE(NKK, 2)
C BACKWARD ELIMINATION PROCEDURE
C
WRITE (6, 3080)
3080 FORMAT(1H1, '******** BACKWARD ELIMINATION PROCEDURE ********')
DO 200 I = 1, NKK
200 KKMOD(I, 1) = 1
NKK1 = NKK - 1
DO 210 MMM = 1, NKK1
NCOU = 0
DO 220 I = 1, NKK
NP = KKMOD(I, 1)
IF (NP .EQ. 0) GO TO 220
NCOU = NCOU + 1
KKMOD(NCOU + 1) = NP
220 CONTINUE
AMAX = 0.0
DO 230 I = 1, NCOU
KEF = KKMOD(I, 3)
KKMOD(I, 3) = 0
KNNN = 0
DO 240 J = 1, NCOU
KEF = KKMOD(J, 3)
IF (KEF .EQ. 0) GO TO 240
KNNN = KNNN + 1
KKMOD(KNNN, 4) = KEF
240 CONTINUE
DO 250 L = 1, KNNN
KP = KKMOD(L, 4)
DO 251 M = 1, KNNN
KQ = KKMOD(M, 4)
AAA(L, M) = CCC(KP, KQ)
251 CONTINUE
AA1(L) = DIFF(KP)
BB1(L) = AA1(L)
250 CONTINUE
CALL SIMEQS(AAA, AA1, KNNN, NCHECZ).
Computer Program of FORBAG

IF (NCHEC2 .EQ. 1) GO TO 161
ASQR=0.0
DO 260 L=1,KNNN
260 ASQR=ASQR+AA1(L)*BR1(L)
IF (ASQR=AMAX) 271,271,270
270 AMAX=ASQR
MHAN=KEE
DO 275 L=1,KNNN
STAND(L)=AA1(L)
KKK(L)=KKMOD(L+4)
275 CONTINUE
271 KKMOD(L+3)=KEE.
230 CONTINUE
WORK(KNNN,1)=AMAX
KKMOD(KNNN+2)=MHAN
STORE(KNNN+1,3)=KKMOD(KNNN,2)
STORE(KNNN+4)=WORK(KNNN,1)
IF (KNNN .NE. 11) GO TO 9601
STORE(1,3)=KKK(1)
9601 CONTINUE
DO 9550 MMP=1,NKK
9550 AA1(MMP)=WORK(MMP,2)
DO 9600 L=1,KNNN
STAND(L)=STAND(L)/SSTAND(LN)
9600 CONTINUE
CALL PRINTA(MMP,MHAN,MA,AMAX,KKK,KKKK,STAND,AA1,2,NKK,YVAR,YMEAN)
KKMOD(MHAN,1)=0
210 CONTINUE
C
C SELECTION PROCEDURES END
C
WRITE(*,4001)
4000 FORMAT(1H1,120X,*** LIST OF ORDERED VARIABLES ***,
1 '1H ,20X,"FORWARD SELECTION PROCEDURE",
2 '1H ,7X,"STEP",5X,"NUMBER",9X,"SS",18X,"SR")
DO 4001 I=1,NKK
NFO=STORE(I,1)
NFOO=KKK(NFO)
FOO=(YVAR*FLOAT(MA-1)-STORE(I,2))/(FLOAT(MA-1)-1.0)
4001 WRITE(*,4002) I,NFOO,STORE(I,2),FOO
4002 FORMAT(1H1,5X,15,5X,15,5X,15,5X,15,5X,7,5X,15,7)
WRITE(*,4003)
4003 FORMAT(1H1,120X,"BACKWARD ELIMINATION PROCEDURE",
1 '1H ,7X,"STEP",5X,"NUMBER",9X,"SS",18X,"SR")
DO 4004 I=1,NKK
NBA=STORE(I,3)
NBAA=KKK(NBA)
BAA=(YVAR*FLOAT(MA-1)-STORE(I,4))/(FLOAT(MA-1)-1.0)
4004 WRITE(*,4002) I,NBAA,STORE(I,4),BAA
WRITE(*,4005)
4005 FORMAT(1H1,120X,"STEP",NUMBER WHICH ORDERS DO NOT COINCIDE",
1 'NROT=0
DO 4010 I=1,NKK
NOM=0
DO 4011 J=1,I
DO 4012 K=1,I
IF (STORE(J,1) .EQ. STORE(K,3)) GO TO 4013
4012 CONTINUE
GO_TO 4011
SUBROUTINE PRINTA (JP, MPVAR, MAP, PMAX, KKP, KPMOD, PSTAND, PA1, MPCOU,  
MPVAR, YPVAR, YPMEAN)
DIMENSION KKP (50), KPMOD (50), PSTAND (50), PA1 (50).
WRITE (6, 3110) JP
NFF=KKP (MPHAN).
IF (MPCOU .EQ. 1) GO TO 9900
WRITE (6, 3121) NFF
WRITE (6, 3130)
MPDF=MAP
JJP=MPVAR-JP
GO TO 9990
WRITE (6, 3120) NFF
WRITE (6, 3130).
JJP=JP
MPDF=MAP+1
9990 CONTINUE
DO 9900 I=1, JJP
MFF=KPMOD (I)
MMFF=KKP (MMFF)
PPPP=PSTAND (I)
WRITE (6, 3140) MMFF, PPPP
9900 CONTINUE
PMEAN1=0.0
DO 9010 I=1, JJP
MFF=KPMOD (I)
PMEAN1=PMEAN1+PSTAND (I)*PA1 (MFF)
9010 CONTINUE
PMEAN1=YPMEAN-PMEAN1
WRITE (6, 3150) PMEAN1
WRITE (6, 3160) YPMEAN, MPVAR, MAP
WRITE (6, 3190) PMAX
FFFF=YPVAR*FLOAT (MAP)-PMAX)/FLOAT (MAP-JJP-1.0)
WRITE (6, 3200) FFFF
STANDD=SORT (FFFF)
WRITE (6, 3220) STANDD
MMDF=MAP-JJP-1
CORRE=SORT (PMA_
YVAR*FLOAT (MAP-1))
WRITE (6, 3230) CORRE
WRITE (6, 3240) JJP, MMDF
3110 FORMAT (1H10X, 'STEP ( \^, I3, \^), \^, )')
3120 FORMAT (1H10X, 'ENTERING INDEPENDENT VARIABLE NUMBER . . . X( \^,  
1 \^, I5, \^), )')
3121 FORMAT (1H10X, 'EXCLUDING INDEPENDENT VARIABLE NUMBER . . . X( \^,  
1 \^, I5, \^), )')
3130 FORMAT (1H10X, 'REGRESSION COEFFICIENTS\^, )')
Computer Program of FORBAG

3140 FORMAT (IH, 24X, ' ( ', 15, ' ) = ', E15.7)
3150 FORMAT (IH, 22X, 'CONSTANT', 10X, E15.7, ' / ')
3160 FORMAT (IH, 20X, 'MEAN VALUE OF DEPENDENT VARIABLE = ', E15.7)
   1 /IH, 20X, 'UNBIASED VARIANCE OF DEPENDENT VARIABLE = ', E15.7
   2 /IH, 20X, 'NUMBER OF DATA = ', E11.7
3190 FORMAT (IH, 20X, 'SUM OF SQUARES DUE TO REGRESSION (SS) = ', E15.7)
3200 FORMAT (IH, 20X, 'MEAN SQUARE DUE TO RESIDUAL VARIATION (SR) = ', E15.7)
3220 FORMAT (IH, 20X, 'SQUARE ROOT OF SR = ', E15.7)
3230 FORMAT (IH, 20X, 'MULTIPLE CORRELATION COEFFICIENT (R) = ', E15.7)
3240 FORMAT (IH, 20X, 'DEGREE OF FREEDOM ( ', 14, ' , ', 17, ' ) = ', E15.7, ' / ')
RETURN
END

SUBROUTINE SIMEQS (BBR, DD, NDim, NCHECK)
DIMENSION BBR(50, 50), DD(50)
NCHECK = 0
DO 10 K = 1, NDim
   P = BBR(K, K)
   IF (P .EQ. 0.0) GO TO 100
   K1 = K + 1.
   IF (K1 .GT. NDim) GO TO 21
   DO 20 J = K1, NDim
      20 BBR(K, J) = BBR(K, J)/P
      DD(K) = DD(K)/P
   DO 30 I = 1, NDim
      30 P = BBR(I, K)
      IF (K1 .GT. NDim) GO TO 41
      DO 40 J = K1, NDim
         40 BBR(I, J) = BBR(I, J) - BBR(K, J)*P
      DO 41 CONTINUE
   41 CONTINUE
   CONTINUE
100 NCHECK = 1
RETURN
END
Table 2, Data

Variance covariance matrix (AAA)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$aa_{11}$</td>
<td>226.314</td>
</tr>
<tr>
<td>$aa_{22}$</td>
<td>242.141</td>
</tr>
<tr>
<td>$aa_{33}$</td>
<td>41.026</td>
</tr>
<tr>
<td>$aa_{44}$</td>
<td>3.167</td>
</tr>
<tr>
<td>$aa_{55}$</td>
<td>280.167</td>
</tr>
</tbody>
</table>

Mean value (AA1)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c_1$</td>
<td>95.423</td>
</tr>
<tr>
<td>$c_2$</td>
<td>7.462</td>
</tr>
<tr>
<td>$c_3$</td>
<td>48.154</td>
</tr>
<tr>
<td>$c_4$</td>
<td>11.769</td>
</tr>
<tr>
<td>$c_5$</td>
<td>30.000</td>
</tr>
</tbody>
</table>

Dependent and independent variables number (KKK)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_1$</td>
<td>5</td>
</tr>
<tr>
<td>$k_2$</td>
<td>1</td>
</tr>
<tr>
<td>$k_3$</td>
<td>2</td>
</tr>
<tr>
<td>$k_4$</td>
<td>3</td>
</tr>
<tr>
<td>$k_5$</td>
<td>4</td>
</tr>
</tbody>
</table>

$n = 13$ (MA)

$k = 4$ (NKK : $k + 1$)
Table 3, Computer Output

<table>
<thead>
<tr>
<th>DEPENDENT VARIABLE NUMBER</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>INDEPENDENT VARIABLE NUMBER</td>
<td>1</td>
</tr>
</tbody>
</table>

NORMALIZED COEFFICIENTS

<table>
<thead>
<tr>
<th>INDEPENDENT VARIABLE NUMBER</th>
<th>NORMALIZED COEFFICIENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.807987E+02</td>
</tr>
<tr>
<td>2</td>
<td>4.25736E+02</td>
</tr>
<tr>
<td>3</td>
<td>-2.786328E+02</td>
</tr>
<tr>
<td>4</td>
<td>-4.280066E+02</td>
</tr>
</tbody>
</table>

CORRELATION MATRIX

<table>
<thead>
<tr>
<th>BETWEEN_INDEPENDENT_VARIABLES</th>
<th>CORRELATION_COEFFICIENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1, 1)</td>
<td>1.000000E+01</td>
</tr>
<tr>
<td>(1, 2)</td>
<td>2.285795E+00</td>
</tr>
<tr>
<td>(1, 3)</td>
<td>-8.241338E+00</td>
</tr>
<tr>
<td>(1, 4)</td>
<td>-4.254651E+00</td>
</tr>
<tr>
<td>(2, 2)</td>
<td>1.000000E+01</td>
</tr>
<tr>
<td>(2, 3)</td>
<td>-1.392424E+00</td>
</tr>
<tr>
<td>(2, 4)</td>
<td>-9.729550E+00</td>
</tr>
<tr>
<td>(3, 3)</td>
<td>1.000000E+01</td>
</tr>
<tr>
<td>(3, 4)</td>
<td>-2.953700E-01</td>
</tr>
<tr>
<td>(4, 4)</td>
<td>-1.000000E+01</td>
</tr>
</tbody>
</table>
** FORWARD SELECTION PROCEDURE **

* STEP ( 1 )

ENTERING INDEPENDENT VARIABLE NUMBER • • • X ( 4 )

REGRESSION COEFFICIENTS

\[ B ( 4 ) = -7.381618E+00 \]
\[ \text{CONSTANT} = 1.175679E+03 \]

MEAN VALUE OF DEPENDENT VARIABLE = \( 9.542308E+02 \)
UNBIASED VARIANCE OF DEPENDENT VARIABLE = \( 2.263136E+03 \)
NUMBER OF DATA = 13

SUM OF SQUARES DUE TO REGRESSION (SS) = \( 1.831826E+04 \)
MEAN SQUARE DUE TO RESIDUAL VARIATION (SR) = \( 8.035154E+02 \)
SQUARE ROOT OF SR = \( 8.963902E+01 \)
MULTIPLE CORRELATION COEFFICIENT (R) = \( 8.213050E+00 \)

DEGREE OF FREEDOM ( 1, 11 )

* STEP ( 2 )

ENTERING INDEPENDENT VARIABLE NUMBER • • • X ( 1 )

REGRESSION COEFFICIENTS

\[ B ( 4 ) = -6.139536E+00 \]
\[ B ( 1 ) = 1.439958E+01 \]
\[ \text{CONSTANT} = 1.030974E+03 \]

MEAN VALUE OF DEPENDENT VARIABLE = \( 9.542308E+02 \)
UNBIASED VARIANCE OF DEPENDENT VARIABLE = \( 2.263136E+03 \)
NUMBER OF DATA = 13

SUM OF SQUARES DUE TO REGRESSION (SS) = \( 2.641001E+04 \)
MEAN SQUARE DUE TO RESIDUAL VARIATION (SR) = \( 7.476213E+01 \)
SQUARE ROOT OF SR = \( 7.342671E+00 \)
MULTIPLE CORRELATION COEFFICIENT (R) = \( 9.864395E+00 \)

DEGREE OF FREEDOM ( 2, 10 )

\[
\begin{pmatrix}
N_B, \text{in step 2, } I_2 = 1 \\
\hat{Y} = -0.6140x_4 + 1.4400x_1 + 103.0974 \\
S_{y2}^2 = 7.4762, \ D(4,1) = 2641.001
\end{pmatrix}
\]
*STEP (3)*

ENTERING INDEPENDENT VARIABLE NUMBER • • • x( 2 )

REGRESSION COEFFICIENTS

\[
\begin{align*}
B \ ( \ 4 \ ) &= \ -2.36502 \times 10^0 \\
B \ ( \ 1 \ ) &= \ 1.451938 \times 10^1 \\
B \ ( \ 2 \ ) &= \ 4.161098 \times 10^0 \\
\text{CONSTANT} &= \ 7.164830 \times 10^0 \\
\end{align*}
\]

MEAN VALUE OF DEPENDENT VARIABLE = \ 9.542308 \times 10^2

UNBIASED VARIANCE OF DEPENDENT VARIABLE = \ 2.263136 \times 10^3

NUMBER OF DATA = \ 13

SUM OF SQUARES DUE TO REGRESSION (SS) = \ 2.667790 \times 10^4

MEAN SQUARE DUE TO RESIDUAL VARIATION (SR) = \ 5.330305 \times 10^1

SQUARE ROOT OF SR = \ 2.308745 \times 10^1

MULTIPLE CORRELATION COEFFICIENT (R) = \ 9.911286 \times 10^0

DEGREE OF FREEDOM ( . . . 3 , . . . 9 )

*STEP (4)*

ENTERING INDEPENDENT VARIABLE NUMBER • • • x( 3 )

REGRESSION COEFFICIENTS

\[
\begin{align*}
B \ ( \ 4 \ ) &= \ -1.440606 \times 10^0 \\
B \ ( \ 1 \ ) &= \ 1.551103 \times 10^1 \\
B \ ( \ 2 \ ) &= \ 5.101681 \times 10^0 \\
B \ ( \ 3 \ ) &= \ 1.019099 \times 10^0 \\
\text{CONSTANT} &= \ 6.240532 \times 10^2 \\
\end{align*}
\]

MEAN VALUE OF DEPENDENT VARIABLE = \ 9.542308 \times 10^2

UNBIASED VARIANCE OF DEPENDENT VARIABLE = \ 2.263136 \times 10^3

NUMBER OF DATA = \ 13

SUM OF SQUARES DUE TO REGRESSION (SS) = \ 2.667899 \times 10^4

MEAN SQUARE DUE TO RESIDUAL VARIATION (SR) = \ 5.982957 \times 10^1

SQUARE ROOT OF SR = \ 2.446008 \times 10^1

MULTIPLE CORRELATION COEFFICIENT (R) = \ 9.911486 \times 10^0

DEGREE OF FREEDOM ( . . . 4 , . . . 8 )
** BACKWARD ELIMINATION PROCEDURE **

* STEP ( 1 )

EXCLUDING INDEPENDENT VARIABLE NUMBER • • • X(     3 )

REGRESSION COEFFICIENTS

\[
\begin{align*}
B ( 1 ) &= 1.451938E+01 \\
B ( 2 ) &= 4.161098E+00 \\
B ( 4 ) &= -2365402E+00 \\
\text{CONSTANT} &= 7164830E+02 \\
\end{align*}
\]

- MEAN VALUE OF DEPENDENT VARIABLE = \*9542308E+02
- UNBIASED VARIANCE OF DEPENDENT VARIABLE = \*2263136E+03
- NUMBER OF DATA = 13
- SUM OF SQUARES DUE TO REGRESSION (SS) = \*2667790E+04
- MEAN SQUARE DUE TO RESIDUAL VARIATION (SR) = \*5330305E+01
- SQUARE ROOT OF SR = \*2308745E+01
- MULTIPLE CORRELATION COEFFICIENT (R) = \*9911284E+00

DEGREE OF FREEDOM ( 3 , 9 )

* STEP ( 2 )

EXCLUDING INDEPENDENT VARIABLE NUMBER • • • X (     4 )

REGRESSION COEFFICIENTS

\[
\begin{align*}
B ( 1 ) &= 1.468306E+01 \\
B ( 2 ) &= 6622505E+00 \\
\text{CONSTANT} &= 5257735E+02 \\
\end{align*}
\]

- MEAN VALUE OF DEPENDENT VARIABLE = \*9542308E+02
- UNBIASED VARIANCE OF DEPENDENT VARIABLE = \*2263136E+03
- NUMBER OF DATA = 13
- SUM OF SQUARES DUE TO REGRESSION (SS) = \*2657859E+04
- MEAN SQUARE DUE TO RESIDUAL VARIATION (SR) = \*5790450E+01
- SQUARE ROOT OF SR = \*2406335E+01
- MULTIPLE CORRELATION COEFFICIENT (R) = \*9892817E+00

DEGREE OF FREEDOM ( 2 , 10 )

N.B. in step 2, \( J_2 = 4 \)

\[
\begin{pmatrix}
\hat{Y} = 1.4683x_1 + 0.6623x_2 + 52.5774 \\
\end{pmatrix}
\]

\[
\begin{align*}
\hat{\sigma}^2 &= 5.7905 \\
\sigma^2 &= 2657.859 \\
\end{align*}
\]
Computer Program of FORBAG

* STEP (3)

EXCLUDING INDEPENDENT VARIABLE NUMBER ••• X_1

REGRESSION COEFFICIENTS

B (2) = 7891248E+00
CONSTANT = 5742368E+02

MEAN VALUE OF DEPENDENT VARIABLE = 9542308E+02
UNBIASED VARIANCE OF DEPENDENT VARIABLE = 2263136E+03
NUMBER OF DATA = 13

SUM OF SQUARES DUE TO REGRESSION (SS) = 1809427E+04
MEAN SQUARE DUE TO RESIDUAL VARIATION (SR) = 8239421E+02
SQUARE ROOT OF SR = 9077126E+01
MULTIPLE CORRELATION COEFFICIENT (R) = 8162526E+00

DEGREE OF FREEDOM (11)

(N.B. in step 3, J_1 = 2)

*** LIST OF ORDERED VARIABLES ***

FORWARD SELECTION PROCEDURE

<table>
<thead>
<tr>
<th>STEP</th>
<th>NUMBER</th>
<th>SS</th>
<th>SR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>1.831896E+04</td>
<td>8.035154E+02</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2.667990E+04</td>
<td>5.330305E+01</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2.667990E+04</td>
<td>5.982957E+01</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>2.667990E+04</td>
<td>5.982957E+01</td>
</tr>
</tbody>
</table>

BACKWARD ELIMINATION PROCEDURE

<table>
<thead>
<tr>
<th>STEP</th>
<th>NUMBER</th>
<th>SS</th>
<th>SR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>1.809427E+04</td>
<td>8.239421E+02</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2.657859E+04</td>
<td>5.790450E+01</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>2.667990E+04</td>
<td>5.982957E+01</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>2.667990E+04</td>
<td>5.982957E+01</td>
</tr>
</tbody>
</table>

(N.B. eliminated variables are shown by the
order of J_1, J_2, ..., J_k + SS = D(J_1, J_2, ..., J_m))

STEP NUMBER WHICH ORDERS DO NOT COINCIDE

(N.B. if (J_1, J_2, ..., J_m) ≠(J_1, J_2, ..., J_m),)

1 (step number m is printed out)